

Statistical models for dynamics in extreme value processes

Bernhard Spangl¹, Sascha Desmettre², Peter Ruckdeschel³

¹ Institute of Applied Statistics and Computing, BOKU, Vienna, Austria

² Department of Mathematics, TU Kaiserslautern, Germany

³ Institute for Mathematics, University Oldenburg, Germany

E-mail for correspondence: bernhard.spangl@boku.ac.at

Abstract: We study four different approaches to model time-dependent extremal behavior: dynamics introduced by (a) a state-space model (SSM), (b) a shot-noise-type process with GPD marginals, (c) a copula-based autoregressive model with GPD marginals, and (d) a GLM with GPD marginals (and previous extremal events as regressors). Each of the models is fit against data, and from the fitted data, we simulate corresponding paths according to the respective fitted models. At this simulated data, the respective dependence structure is analyzed in copula plots and judged against its capacity to fit the corresponding inter-arrival distribution.

Keywords: time-dependence; state-space model; GLM; shot-noise-type process; extreme values.

1 Motivation and issues

A challenge in dealing with extreme events in river discharge data is to capture well time dynamics of these extremes, in particular in the presence of seasonal effects and trends.

We will provide models which are able to capture the extreme behaviour and provide simple and parsimonious, but yet flexible dynamics.

The goal is to be able to assess the magnitude of extreme events as well as the inter-arrival time distribution of those extremes by simulation (see also Khaliq et al., 2006).

2 Data generating processes

To address these issues, we discuss four different approaches to incorporate dynamics into extreme value processes. For different choices of parameters, each of these approaches is illustrated by typical realizations to assess the induced dynamics and by lagged PP plots to grasp the respective dependence structures.

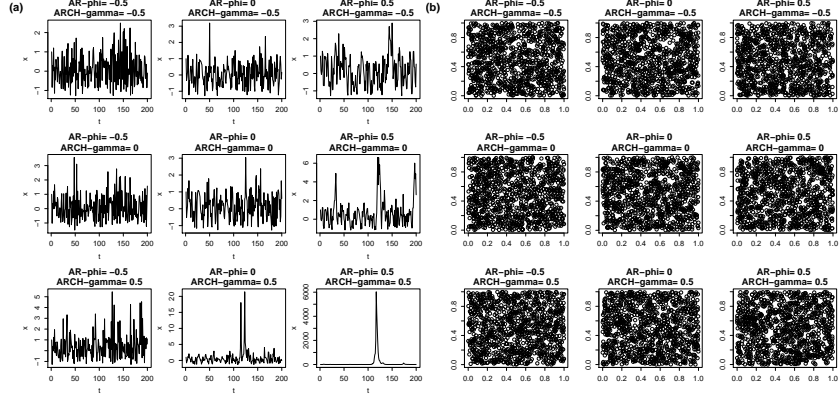


FIGURE 1. SSM approach: (a) typical realizations, and (b) lagged PP plots.

2.1 SSM approach

As a first approach we propose to use the following model:

$$X_t = \mu_t + \sigma_t v_t ,$$

with

$$\begin{aligned} \mu_t &= \mu_0 + \varphi X_{t-1} , \\ \log \sigma_t &= \log \sigma_0 + \gamma \log(1 + (X_{t-1} - \mu_{t-1})^2) . \end{aligned}$$

This model is comparable to an AR-EARCH one and can be easily extended to an ARMA-EGARCH model or other even more complex ones.

The iid innovations v_t are generated according to the following two-step procedure. First, we choose $v'_t \sim F_0$, where, e.g., $F_0 = \mathcal{N}(0, 1)$. Then, given a certain threshold τ the innovations are enriched by $v''_t \sim \text{GPD}$, i.e.,

$$v_t = \begin{cases} v'_t & : |v'_t| < \tau , \\ v''_t \sim \text{GPD}(\mu = 0, \sigma = 1, \xi) & : \text{else} . \end{cases}$$

ξ is chosen according to the maximum domain of attraction of F_0 .

This state-space model approach allows for separate estimation of time-dependency (via filtering) and of GPD-parameters, μ , σ , and ξ , of the marginal distribution of the innovations. Typical realizations are shown in Figure 1, together with the corresponding lagged PP plots.

2.2 Shot-noise-type approach

As a second approach we suggest to use the following shot-noise-type process:

$$X_t = U_t Y_t + (1 - U_t) \min\{v_t, Y_t\} ,$$

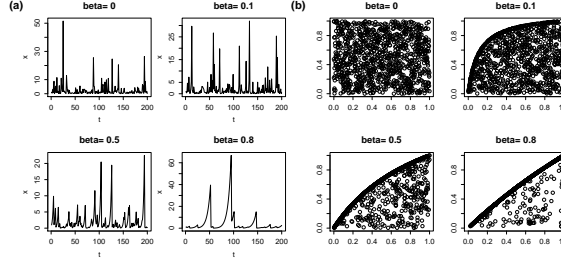


FIGURE 2. Shot-noise-type approach: (a) typical realizations, and (b) lagged PP plots.

where $U_t \sim \text{Bin}(0, \beta)$ and $v_t \sim G = \text{GPD}(\mu, \sigma, \xi)$. Let G and G^{-1} also denote the corresponding cdf and quantile function, respectively. Moreover, $\{U_t\}_t$ and $\{v_t\}_t$ are stochastically independent, and (U_t, v_t) is independent of $\{X_s\}_{s < t}$. Then $Y_t = f_\beta(X_{t-1})$ with

$$\begin{aligned} f_\beta &= G^{-1} \circ f_\beta^0 \circ G, \\ f_\beta^0 &= u / [(1-u)\beta + u]. \end{aligned}$$

Hence, $X_t \sim G$. The parameter $\beta \in [0, 1]$ controls the dependency. Again, this approach allows for separate estimation of time-dependency (number of dropdowns) and of GPD-parameters, μ, σ, ξ , of the marginal distribution. Figure 2 shows typical realizations of the shot-noise-type (SNT) approach, together with the corresponding lagged PP plots. We refer to Desmettre et al. (2015) for the definition and theoretical foundations of this process as well as its application in liquidity risk management.

2.3 Gaussian-copula approach

Next, we propose to use a copula-based autoregressive model defined by

$$X_t = G^{-1} \circ F \circ Y_t,$$

with

$$Y_t = \rho Y_{t-1} + \bar{\rho} v_t,$$

where $\bar{\rho}^2 + \rho^2 = 1$, and $v_t \stackrel{\text{iid}}{\sim} F = \mathcal{N}(0, 1)$. Again, let G and G^{-1} denote the cdf and quantile function of the $\text{GPD}(\mu, \sigma, \xi)$. Hence, $X_t \sim G$, and the parameter $\rho \in [-1, 1]$ controls the dependency. In the same way we may transform a more complex model, e.g., a Gaussian ARMA-model to a GPD-process. This again separates the estimation of time-dependency (Gaussian ARMA) and of GPD-parameters, μ, σ, ξ , of the marginal distribution. In Figure 3 we see typical realizations of the Gaussian-copula (GC) approach, together with the corresponding lagged PP plots.

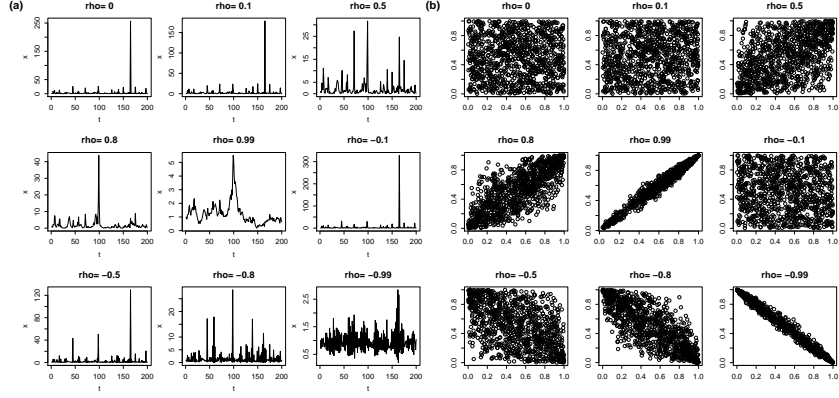


FIGURE 3. Gaussian-copula approach: (a) typical realizations, and (b) lagged PP plots.

2.4 GLM approach

Last, we suggest a parameter driven approach. We define our GPD-process by

$$X_t \sim \text{GPD}(\mu, \sigma_t, \xi_t) .$$

The scale parameter σ_t is given by

$$\sigma_t = \ell_1(\beta_1 + \beta_2 h_1(X_{t-1})) ,$$

with ℓ_1 and h_1 properly chosen, e.g., $\ell_1: \mathbb{R} \rightarrow \mathbb{R}_+$, and $h_1(x) = x^2$.

The shape parameter ξ_t is given by

$$\xi_t = \ell_2(\gamma_1 + \gamma_2 h_2(X_{t-1})) ,$$

with ℓ_2 and h_2 again properly chosen, e.g., $\ell_2: \mathbb{R} \rightarrow (-0.5, 2.5)$, and $h_2(x) = \log(1 + |x|)$. h_2 controls the tails. Moreover, we note that it is essentially to chose the codomain of ℓ_2 equal to $(-0.5, 2.5)$.

Typical realizations of the GLM approach are plotted in Figure 4. Here, lagged PP plots do not make sense as the dependence structure is modeled by fitting the GPD-parameters by GLMs where previous observations enter as regressors. See Pupashenko et al. (2014) for theoretical foundations.

3 Evaluation and real world data set

To evaluate our four approaches we use daily average discharge data of the Danube river at Donauwörth from 1978 to 2008. We obtain corresponding model fits and simulated data as visible in the following pictures.

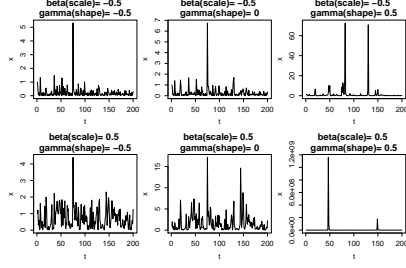


FIGURE 4. GLM approach: typical realizations.

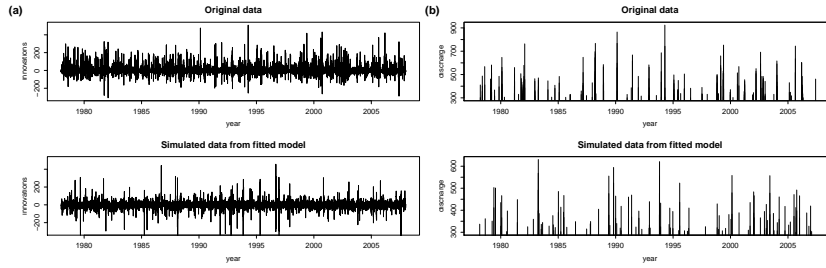


FIGURE 5. Real data vs. simulation: (a) SSM approach, and (b) shot-noise-type approach.

In Figures 5 and 6 the upper panels always show the original data and the lower panels one simulated path.

Figure 5a shows the innovations of the original series after filtering and the simulated innovations using the SSM approach. Here, instead of using an EARCH model as proposed in Section 2.1, we use a GARCH(1,1) model with t -distributed innovations. Comparing the two series in Figure 5a we see that in the lower panel the volatility clusters are reproduced quite well. However, negative innovations are over-represented in the simulated series. In Figure 5b only the extreme events above a threshold of 300 of the original data as well as the simulated ones using the shot-noise-type approach are plotted. Comparing the two plots in Figure 5b we note that the distribution of extreme events, i.e., the inter-arrival time distribution of extremes, coincide well, whereas the level of discharge is slightly under-estimated by the simulated process.

In the upper panels of Figure 6 the original detrended and deseasonalized series is displayed. The lower panel of Figure 6a shows one simulated path using the Gaussian-copula approach whereas in Figure 6b we see a simulated path using the GLM approach. Comparing the lower panels of Figures 6a and 6b with the upper ones we see that the level of discharge is estimated very well by the Gaussian-copula and the GLM approach. How-

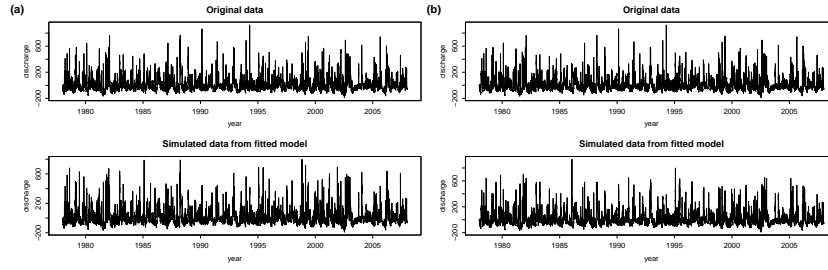


FIGURE 6. Real data vs. simulation: (a) Gaussian-copula approach, and (b) GLM approach.

ever, we note that using these two approaches we are not able to assess the inter-arrival times.

4 Summary and conclusion

We presented four flexible and parametric approaches to model dynamics as well as extremes. We remark that all models proposed are parsimonious ones. Moreover, we are able to get grip on the inter-arrival time distribution of extremes. The SSM and GLM approaches are able to model complex dynamics. The SNT as well as the SSM approach are able to assess the inter-arrival time distribution of the extreme events well. The GC approach can easily be extended to capture multivariate dependency.

Acknowledgments: All authors gratefully acknowledge financial support by the Volkswagen Foundation for the project “Robust Risk Estimation”, <http://www.mathematik.uni-kl.de/~wwwfm/RobustRiskEstimation>.

References

- Desmettre, S., de Kock, J., Ruckdeschel, P. and Seifried, F.T. (2015). Generalized Pareto Processes and Liquidity. *Working Paper*.
- Khaliq, M.N., Ouarda, T.B.M.J., Ondo, J.-C., Gachon, P., and Bobee, B. (2006). Frequency analysis of a sequence of dependent and/or non-stationary hydro-meteorological observations: A review. *Journal of Hydrology*, **329**, 534–552.
- Pupashenko, D., Ruckdeschel, P., and Kohl, M. (2015). L_2 Differentiability of Generalized Linear Models. *Statist. Probab. Lett.* **97**: 155–164.